Documenting provenance of science in a state government agency

Matthew W Miles [a], **Angela London** [a], **Robin Green** [a], **Amy Teitzel** [a], **Jikke W. Adema** [a]

[a] *Department of Environment, Water and Natural Resources. South Australia*
Email: matthew.miles@sa.gov.au

Abstract: The South Australian Department of Environment Water and Natural Resources (DEWNR) has developed a comprehensive suite of standards, tools and guidelines to improve the quality and transparency of the science we produce. The approach is named MEK (Managing Environmental Knowledge). The MEK suite of resources supports the department's Information Management Framework and is aligned to South Australia's Digital by Default and Open Data agendas. DEWNRs Project Management Framework (PMF) does not currently address data management in projects. The MEK fills this gap by adding steps to the project phases which enable data supply chain management. This reminds projects of the importance of early data management planning and which standards or guidelines to refer to during the execution and delivery of a project. MEK tools support project managers and scientists to explicitly document data supply chains so that provenance of intermediate data outputs and publishable products is clear and accessible. The MEK aims to (a) improve the security, integrity and availability of data produced through projects (b) establish provenance, ensure transparency and enable re-use of data products and (c) enable effective use of existing enterprise information systems.

MEK resources are made available to employees through the department's intranet. The tools include (i) **data planning form** for estimating resources and broad needs of a project in relation to data management, (ii) **data charts** that provide a visual way of describing data supply chains and (iii) **data catalogue** for storing detailed metadata of each element in the data charts. The guidelines that underpin these tools include: **data storage** describing how to make use of the various corporate systems and applications, **evaluation** detailing the peer review procedures for major deliverables and evaluation guidelines for other project outputs, **data handling** describing information classification and sensitive data handling, and **publication** including proofing and publishing procedures.

Government agencies which rely on data to produce scientific outputs require a synergy of people, process and technologies to be effective and achieve strategic objectives. In DEWNR, an organisation which was recently formed as an amalgam of a number of departments, a disparate mix of platforms, applications and procedures have evolved over time. The result of this is that there is lack of consistency across the organisation in data management processes.

The South Australian Government's Digital by Default, Open Data and Data Sharing agendas have contributed to the need to better understand the provenance of DEWNR's science information. Simply printing a run of documents is no longer sufficient, we must openly provide intermediate datasets and products such as models, as well as communications material. Therefore, clear provenance is critical to ensure outputs provided to downstream users are evaluated and authorised. The MEK resources are enabling the department to improve its culture around data management, deliver improved science outputs and thereby enable evidence based decision making to support the management of our State's natural resources.

Keywords: *Knowledge management, data management, provenance*

1.   INTRODUCTION

Changing organisational structures coupled with rapidly changing technologies has resulted in unsystematic and disjointed practices in data management within DEWNR. In 2012, a single agency was formed that amalgamated functions relating to environment and heritage, water and natural resource management regions. The new agency, DEWNR, was challenged to operate as a unified body after a number of years operating as separate functions under separate governance regimes. The need for consistency of data management processes was a known issue however organisational cultural differences made this a challenging outcome to achieve. The volume and variety of data coupled with inconsistent data management practices meant the goals of the departments Information Management Framework were a long way off in relation to science outputs.

While the department has a corporate Project Management Framework (PMF) it doesn't currently provide direction in regards to data management. The provenance of science project outputs was at best captured cursorily within deliverable reports or metadata statements and at worst maintained in project officers' minds or notes.

The MEK (Managing Environmental Knowledge) Procedure inserts steps into each PMF phase (summarised as MEK Project Guide in Figure 1) to ensure data management is considered in the early stages of project planning and directs projects to data standards or guidelines to refer to during delivery of a project. MEK resources (DEWNR 2017) support project managers and scientists to explicitly plan and document data supply chains so that provenance of intermediate data outputs and publishable products is clear. This aims to a) improve the quality and dependability of the science produced through ensuring outputs are evaluated and authorised according to organisational standards b) ensure transparency and reusability of data products and c) enable effective use of existing enterprise information systems.
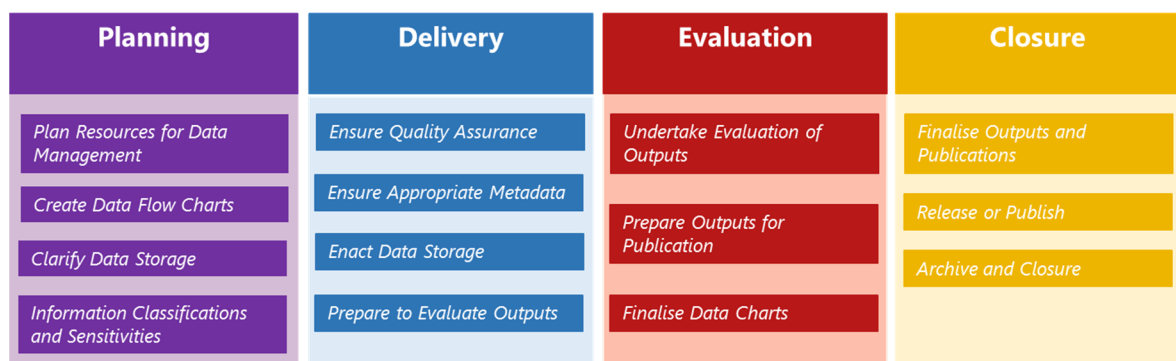

Figure 1. The MEK Project Guide : explicit data and information management steps within project management phases

2.   WHY DID WE NEED MEK?

The organisational structure of DEWNR has given rise to varying data management processes across the department. A lack of compliance with information policies and incomplete focus on the value of metadata has resulted in a "collect-once, use-once" mentality in some projects and programs in relation to data assets. At any time, regions and branches could be concurrently undertaking hundreds of projects some of which form part of larger programs of work which run over many years. While some experienced staff use enterprise information systems for their work, some project data has been stored in a variety of repositories many of which are not accessible outside of the originating business unit (e.g. local network folders). Additionally, data management has historically been seen as an overhead that takes time away from delivering a project.

While some programs or business units have developed systems and processes to effectively record and deliver their outputs, the lack of visibility, discoverability and accessibility of some data being generated within projects has impacted DEWNR's ability to share, innovate, reuse, collaborate and learn from our collective knowledge. Prior to MEK, some projects did periodically tidy and store key datasets if and when an officer or manager with data management interests was involved. The MEK procedure and tools were needed to promote this to business-as-usual and get staff to understand proper data management and how it adds value to the agency.

MEK tools help deliver better data management throughout projects and enable data products to be accessible beyond the life of the project. The tools and guidelines assist with:

- staff in being confident they're using the most recent data and information available

- how knowledge can be used and re-used: "collect once use many times"

- ensuring knowledge won't be lost: its context will be clear to anyone that wants to use it now and in the future

- staff recognise that their project's knowledge is a small but important cog in a bigger system

- project data, previously held in silos, becoming discoverable and available for re-use.

- Avoiding staff accidentally get in trouble for releasing the wrong information

It's important to note that MEK resources were designed to augment and work with frameworks and guidelines that had developed at a corporate level including the Information Management Framework, Information Classification Guidelines, Information Licensing Guidelines and Information Management Roles and Responsibilities.

3. MEK TOOLS

These are a series of templates for planning and documenting a standard set of activities during project management phases. A fact sheet summarising all MEK resources is available on-line (DEWNR 2017b).

1. **MEK Planning Template**[1] assists resource planning for data management related to the size and topics of the project. This is completed during the planning phase of a project to identify project roles, determine project storage requirements, and broadly describe data inputs, outputs and audience. The Planning Template enables project managers to identify the level of resourcing needed in relation to data management and publishing.

2. **MEK Flowcharts**[2] are designed to illustrate project workflows. They describe data flows and supply chains including value adding steps and provide visual representation of classification, licensing, evaluation, approvals and authorisation. The structure of these charts matches in with project management phases and enables project data to be displayed as either working data (in blue icons) evaluated outputs (red icons) or authorised products (gold icons). Operationally this is the jewel in the MEK crown. It is instrumental in forming and guiding detailed conversation about data flows, enabling clear language and understanding to be shared by all stakeholders. The MEK procedure requires flowcharts to be started in the planning phase, refined in delivery phase and finalised in the evaluation phase.

3. **MEK Catalogue Template**[3] is used to document metadata for all project data and information. Elements in the Flowcharts are given identifiers that match more detailed metadata records in the catalogue

4. **MEK Authorisation Checklist**[4] provides a means of ensuring products are ready for publication.

5. **MEK Closure Checklist**[5] completes the MEK procedure. It ensures the important components of data and information management have been completed.

---

[1] https://data.environment.sa.gov.au/Content/Publications/MEK-Planning-Template.pdf
[2] https://data.environment.sa.gov.au/Content/Publications/MEK-Flowcharts.pdf
[3] https://data.environment.sa.gov.au/Content/Publications/MEK-Catalogue-Template.pdf
[4] https://data.environment.sa.gov.au/Content/Publications/MEK-Authorisation-Checklist.pdf
[5] https://data.environment.sa.gov.au/Content/Publications/MEK-Closure-Checklist.pdf

BRUVS Monitoring Dataflow
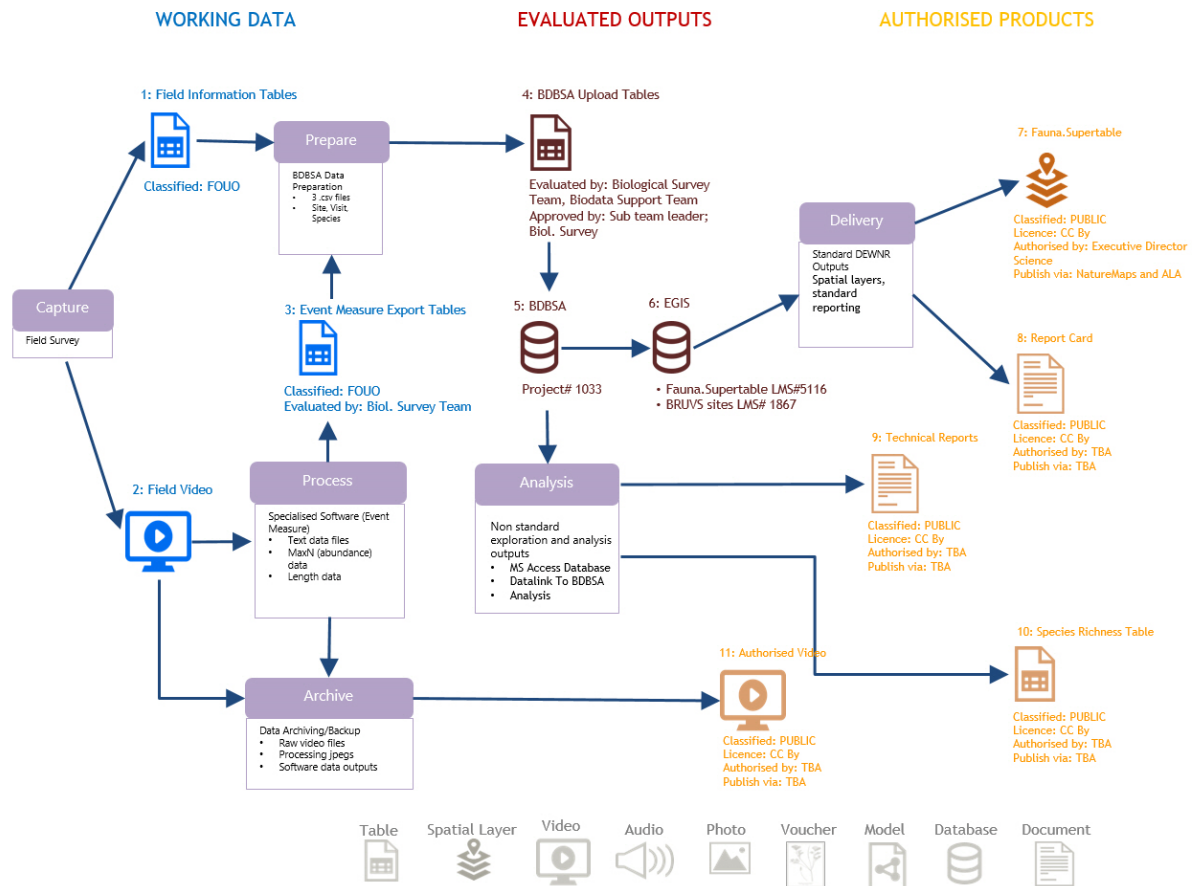## Baited remote underwater video system

Figure 2. Example of data chart showing BRUVS data processing steps.

4.   MEK GUIDELINES

MEK guidelines describe details relating to data management and processing standards within DEWNR. These enable project officers to understand prevailing approaches to corporate data storage, data handling and publication. A fact sheet summarising all MEK resources is available on-line (DEWNR 2017b)

1. Storage Guidelines provide a summary of storage options including iShare, structured databases and delivery channels. The guidelines include how to make use of the various corporate data systems and applications.

2. Evaluation Guidelines describe an evaluation framework used to assess the quality of data and information that is collected, analysed and disseminated to DEWNR science products. Evaluation guidelines, include peer edit review procedures for major deliverables and evaluation principles for spatial and other formats of data.

3. Data Handling Guidelines describe a range of data handling procedures and processes to assist the management of data and information within DEWNR science projects. The guidelines include procedures for handling sensitive data, third party data, data classification and acquisition of data from contracts.

4. Publication Guidelines is a step by step publication process for DEWNR Science products including writing, editing, proofing and publishing procedures.

5. HOW DID WE PRODUCE THE RESOURCES? AND HOW HAVE WE IMPLEMENTED THEM?

A number of commonly occurring problems initiated the development of the MEK suite of resources. Staff had difficulty finding existing project data. Project staff were uncertain about where to store data or how to describe it in a consistent manner. Staff were unaware of what data they could share across the department or externally and third party contracts rarely explicitly addressed the IP of data.

Development started with the idea of decision trees or work flows to guide data management. Looking across the breadth and multi-disciplinary nature of science projects it was clear that creating one explicit procedure from start to finish of a project was unlikely to suffice. In addition, different business areas needed different gaps in information management (IM) filled as they each had existing systems, protocols and procedures for managing data, some more effective, modern and sustainable than others.

The suite of tools and guidelines were designed and created then road tested within a few key projects where data manager roles had been embedded. One region, the SA Murray-Darling Basin (SAMDB) region had also invested in a data mamnagement role (not all business areas did so) and realised the MEK provide a framework to help address a number of IM concerns they had. Feedback from these live test beds enabled improvements to be made and identified possible barriers to uptake.

Refinement of the materials was entrusted to a MEK Working Group which comprised a number of end-users as well as information management specialists. MEK Working Group (with end-users as well as information management specialists as members)

Implementation of the MEK in the SAMDB was approached as a cultural change project. The success of this pilot can be attributed to a number of key factors;

- Sufficient resourcing; a dedicated regional information manager was available to fully support project managers in their familiarisation with, and use of, the MEK tools.

- Executive management buy-in ensured staff were encouraged to take time managing environmental knowledge using the MEK tools.

- Reinforcing the "what's in it for me"/benefits approach was integral to address the ingrained perception that data management takes time away from delivering a project.

- Customising tools to suit the region made the tools relevant and matching existing regional processes, which made it easy for staff to connect with the tools.

- The information manager purposefully captured the current data management practices of each project in the templates during the pilot, as opposed to identifying what should change and capturing the 'best practice' version for each project. This way staff were introduced to the tools and templates in an environment they were familiar with, facilitating understanding of the principles underlying the MEK. This often resulted in project managers themselves identifying data management improvements and short term wins. In the future, the information manager will meet with project managers each year to review the templates and discuss, plan and document improvement opportunities.

A MEK Working Group was formed (including end users, decision makers and information management experts) to review and endorse the resources created, based on the results of the project trials and the experience from the SAMDB pilot. Resources were made available on the agency intranet.

Implementation is now being approached as a cultural change project. Introduction to the vision, the concept of the MEK, as well as the benefits and need for change was communicated from an early stage. The full version of the MEK was introduced as a concept to project and regional staff through a series of workshops. These sessions outline the benefits and provide a practical introduction to the tools as part of project planning and project phases.

Senior management was informed of the benefits of using the MEK, as well as of the fact that resourcing would be required for initial implementation. Using the MEK is mandatory for the Science branch, but not (yet) for most regions and branches. An easy to access support team was established to assist and empower project staff and research staff, to provide training and to remove barriers in using the different tools and guidelines. An initial focus on short-term wins has ensured the quick uptake of some tools and guidelines, which paved the way for the use of the full suite of resources.

To evaluate the success/uptake a scoring rubric was developed by the SAMDB information manager to enable the measurement of data management improvement over time (figure 3). The rubric was used during the SAMDB pilot to set a baseline score for each project and will be used again in 12 months' time to measure the success of the MEK and aligned data management initiatives in achieving behavioural change in the SAMDB. A similar approach is scheduled for the Science branch implementation.

Figure 3. extracts from SAMDB MEK Maturity Assessment Model. Questions relating to Collection/Creation, Storage, Sharing are scored using these rubrics.

6. CONCLUSION

The South Australian Government's Digital by Default, Open Data and Data Sharing agendas (DPC 2017) have contributed to the need to better understand the provenance of DEWNR's science information. Simply printing a run of documents is no longer sufficient, we must openly provide intermediate datasets and products such as models, as well as communications material. Therefore, clear provenance is critical to ensure outputs provided to downstream users are evaluated and authorised. The MEK resources are enabling the department to improve its culture around data management, deliver improved science outputs and thereby enable evidence based decision making to support the management of our State's natural resources.

Government agencies which rely on data to produce scientific outputs require a synergy of people, process and technologies to be effective and achieve strategic objectives. In DEWNR, an organisation which was recently formed as an amalgam of a number of departments, a disparate mix of platforms, applications and procedures have evolved over time. The result of this is that there is lack of consistency across the organisation in data management processes.

The implementation of the MEK tools, templates, checklists and guidelines are expected to produce a number of key benefits:

- make it easier to discover and reuse data regardless of the project from which the data originated;

- eliminate duplication of data collections and data repositories;

- the application of standard, repeatable processes and procedures;

- simplify data sharing and Open Data processes and

- enabling better and faster decision making.

Feedback from pilot projects has been largely positive. Staff have reported that although making data management an explicit part of project management can feel like an additional administrative overhead , the MEK made it significantly easier to find and re-use project data and simplified a number of projects tasks including auditing.

The MEK resources are enabling DEWNR to improve its data culture, and highlight the value of understanding data provenance for enhanced science outputs to deliver transparent evidence for natural resource management decision making, policy, operations and research.

REFERENCES

DEWNR (2017a). Digital Strategy Website. http://digital.env.sa.gov.au/Pages/default.aspx

DEWNR (2017b). MEK Resources fact sheet https://data.environment.sa.gov.au/Content/Publications/MEK-Resources-Factsheet.pdf

DPC (2017). SA Government Digital Government Website. https://digital.sa.gov.au/resources/topic/digital-government